УДК 681.514

ЭМУЛЯЦИЯ ЭКСПЕРТНОГО ПОДХОДА К ОБРАБОТКЕ КРИВЫХ ГДИС НА ОСНОВЕ НЕЧЕТКОГО КРИТЕРИЯ МАКСИМАЛЬНОЙ ИСТИННОСТИ

Борис Изюмов

РГУ нефти и газа им. И.М. Губкина, Россия, 119991, Москва, Ленинский пр-т., 65, тел. (499) 6810102, e-mail: bizyumov@gubkin.ru

Метою гідродинамічних досліджень свердловин (ГДДС) є визначення моделей свердловини, пласта і меж пластів з подальшою ідентифікацію їх параметрів. Традиційно – це завдання покладається на експерта, який зіставляє реальні дані з типовими кривими зміни тиску вручну або за допомогою спеціальних програмних продуктів. При цьому виникає необхідність виявити в даних ДДІВ лінійні ділянки, за якими можливо встановити переважаючий на конкретному часовому проміжку фільтраційний режим (лінійний, радіальний тощо).

Розглядається заснований на непарному критерії максимальної істинності метод кластерного нечіткого регресійного аналізу, що використовує параметри невизначеності вихідних даних у явному вигляді (метод f-регресії) . Розглянутий приклад доводить, що шляхом коректного задання невизначеностей вихідних даних і вибору параметрів методу, традиційності вирішувану експертом завдання установки кількості і параметрів лінійних ділянок кривої можливо вирішити в напівавтоматичному режимі, для подальшої ідентифікації моделі колектора та її параметрів

Ключові слова: емуляція експертного підходу, гідродинамічні дослідження, свердловина, метод *f*-регресії.

Целью анализа данных гидродинамических исследований скважин (ГДИС) является определение моделей скважины, пласта и границ с последующей идентификацией их параметров. Традиционно эта задача возлагается на эксперта, который сопоставляет реальные данные с типичными кривыми изменения давления, вручную или при помощи специальных программных продуктов. При этом возникает необходимость найти в данных ГДИС линейные участки, позволяющие установить преобладающий на конкретном временном промежутке фильтрационный режим (линейный, радиальный и т.д.).

Рассматривается основанный на нечетком критерии максимальной истинности метод кластерного нечеткого регрессионного анализа, использующий параметры неопределенности исходных данных в явном виде (метод f-регрессии). Рассмотренный пример показывает, что путем корректного задания неопределенностей исходных данных и выбора параметров метода традиционную решаемую экспертом задачу установки количества и параметров линейных участков кривой с целью последующей идентификации модели коллектора и ее параметров можно решить в полуавтоматическом режиме.

Ключевые слова: эмуляция экспертного подхода, гидродинамические исследования, скважина, метод *f-регрессии.*

The main aim of well hydrostatic research (WHSR) data analysis is to determine a model of a well, layer and boundaries with the following identification of their parameters. Traditionally this task is performed by an expert who compares real data with typical pressure curves; that is done manually or with a help of special software. Therefore, there arises the task to determine linear areas in WHED data that allows establishing filtration mode dominant on the certain time slot (linear, radial etc.).

This article investigates the method of cluster fuzzy regression analysis which is based on maximum validity fuzzy measure. This method uses parameters of explicit basic data uncertainty (f-regression method). The analyzed example has shown that the task traditionally performed by an expert (set of quantity and parameters of curve linear areas) and can be carried out in semi-automated mode by means of proper set of basic data uncertainty and chooses of method parameters for consequent identification of reservoir model and its parameters.

Keywords: expert approach emulation, hydrodynamic research, well, f-regression method.

Введение. Месторождение является геологическим объектом природного происхождения и содержит заранее неизвестный объем углеводородов, представляющих коммерческий интерес. Современные методы планирования разработки месторождения подразумевают создание трехмерной компьютерной модели месторождения, изначально создаваемой на основе данных сейсморазведки и поисковых скважин и уточняемой по мере поступления данных. Поскольку пробуренные скважины обеспечивают прямой доступ к коллектору, они являются наиболее надежным источником данных. Для извлечения информации, необходимой для оценки месторождения, из (крайне ограниченного) объема данных по скважинам, применяется широкий спектр методов исследования.

Гидродинамические исследования скважин (ГДИС) являются средством оценки ключевых характеристик околоскважинной зоны: скин-фактора, гидропроводности, характеристик геологической формации и т.д. Распространенные методы ГДИС подразумевают эксплуатацию скважины при заданном дебите параллельно с замерами давления на забое скважины. В частности, кривая восстановления давления снимается в течение нескольких часов или десятков часов после остановки скважины (дебит = 0). Профили давления и производной давления по времени отражают несколько одновременно происходящих процессов, слитых в единую кривую. Эффекты, оказываемые этими процессами на результирующую кривую, к настоящему моменту хорошо изучены [1]. Распространенным методом, применяемым инженерами, является сопоставление производной давления по времени в логарифмических координатах с типичными кривыми изменения давления. Линейные участки на профиле производной указывают на преобладание того или иного эффекта, поэтому проведение прямых через группы точек позволяет получить оценку параметров линейных участков рассчитать искомые фильтрационные характеристики.

Целью работы является определение моделей скважины, пласта и границ с последующей идентификацией их параметров.

Результаты. Приводится пример обнаружения линейных участков в данных кривой восстановления давления с использованием доступной информации о неопределенности данных.

Линейный метод *f-регрессии*

Традиционная задача регрессионного анализа состоит в том, чтобы по ограниченному числу наблюдений построить аналитическую модель, наилучшим образом (в определенном смысле) позволяющую предсказывать значения зависимой переменной по значениям независимых переменных. При этом исходные данные часто несовершенны, т.е. содержат неопределенности, обусловленные как неточностью инструментов и методов измерения, так и влиянием неизвестных факторов. Лотфи А. Заде в своей работе [2] подчеркнул важность включения неопределенности исходных данных в постановку аналитических задач, начав с фразы: «Неопределенность является атрибутом информации».

Теория нечетких множеств [3] предоставляет математический аппарат, позволяющий оперировать информацией, содержащей неопределенности различного происхождения. Теория нечетких множеств относится к области прикладных исследований искусственного интеллекта, основополагающим принципом которых является создание методов и систем, использующих те или иные свойства человеческого интеллекта в формализованном виде, для решения широкого спектра задач. Применительно к регрессионному анализу можно сказать о группе методов, способных получать оценки параметров регрессионной модели с учетом априорной неопределенности исходных данных, заданной в явном виде при помощи нечетких чисел. Танака и др. в 1982 г. первыми сформулировали задачу нечеткого регрессионного анализа [4].

Основополагающим понятием здесь является ***нечеткое множество***, описываемое функцией принадлежности.

Определение 1. Пусть U – четкое множество. Пусть далее A – нечеткое множество на U с функцией принадлежности μ: $U \to [0,1]$. Функция принадлежности показывает степень принадлежности элементов множества U нечеткому множеству A. ■

Определение 2. Пусть $U=R^1$ (ось действительных чисел). A называется ***нечетким числом***, если

1. $\exists$ по крайней мере одно число x^* с $\mu_A(x^*) = 1$.
2. μ_A полунепрерывна сверху и квазивыпукла.
3. Все множества α-уровней A_α ограничены для $\alpha \in (0,1]$. ■

Определение 3. Нечеткое число A называется нечетким числом симметричного L-типа (для краткости ***L-числом***), если существуют действительные числа $\bar{a}$, $\alpha(>0)$ и функция L:

1. $L : [0,\infty) \to [0,1]$; $L(0)= 1$, $L(\infty) = 0$.
2. L непрерывна, является строго монотонно убывающей, и функция принадлежности представима в виде

$$\mu_A(t) = L\left(\frac{|t-\bar{a}|}{\alpha}\right). \quad (1)$$

При этом $\bar{a}$ называется средним значением (*модой*) числа, а α – *разбросом* (параметром нечеткости). Формальное представление четких чисел получается при α=0. ■

Предположения метода линейной *f-регрессии,* впервые опубликованного в работе [5], сводятся к следующему:

• Зависимость между входными (независимыми) и выходной (зависимой) переменными описывается четкой параметрической моделью с p параметрами, $a \in R^p$.

• Экспериментальные входные и выходные переменные $X_i = (X_{1i},\ldots,X_{ni})$ и Y_i – нечеткие L-числа с функциями принадлежности $\{\mu_{X_{ji}}\}$ и μ_{Y_i} (здесь $i=1 \div N$ – номер эксперимента, $j=1 \div n$ – номер входной переменной), причем выбрана следующая реализация функции L:

$$L(u) = \exp(-u^m), m > 1, \quad (2)$$

где m – параметр формы нечеткого числа.

Значение $m\approx1$ используется в случае, если среднее известно точно, но о характере разброса известно мало (либо он неизвестен); $m\approx2$ – при высокой уверенности в точности знаний о среднем и разбросе; $m\approx10$ – для чисел интервального типа.

Таким образом, задача ***нечеткой f-регрессии*** [5] состоит в том, чтобы по нечетким экспериментальным данным, неопределенность которых задана в явном виде, определить оптимальный (в определенном смысле) вектор параметров модели a^*.

Мера соответствия между нечеткой точкой и параметрической кривой

Основное и принципиальное отличие постановки задачи нечеткого *f-регрессионного* *анализа* [5], по сравнению с предложенной Танакой и др. [4], является идентификация **четкой** параметрической модели (т.е. модели, параметры которой представлены обычными числами), по **нечетким** входным и выходным экспериментальным данным. Существенным отличием от **статистических методов регрессионного анализа**, основанных на критериях наименьших квадратов или максимального правдоподобия, является предположение о нечетком характере неопределенностей исходных данных, допускающее более широкую трактовку неопределенностей и дающее полный контроль над параметрами этой неопределенности. Отметим, что критика статистических методов сводится к критике ограничений, связанных с предположением о вероятностной природе неопределенностей исходных данных, гипотезе о нормальном распределении ошибок измерения либо с неспособностью методов включать в рассмотрение априорную неопределенность как таковую.

Ключевым понятием метода *f-регрессии* является **нечеткая мера соответствия** между ансамблем нечетких точек и четкой параметрической моделью.

Определение 4 [6]. Пусть $x_i \in \mathrm{R}^n$, $a \in \mathrm{R}^p$, и пусть $y = f(x,a)$ – четкая функция. Для нечеткой точки Q_i определим ее *меру соответствия* функции f (для заданного вектора a) как

$$M_i(\boldsymbol{a}) = \sup_{\boldsymbol{x}\in\mathbf{R}^n} \mu_{\boldsymbol{Q}_i}\left(\boldsymbol{x}, f(\boldsymbol{x},\boldsymbol{a})\right). \qquad (3)$$

Можно рассматривать $M_i\,(a)$ как меру истинности утверждения «нечеткая точка Q_i лежит на кривой f с параметрами a», причем при $M_i\,(a)$=1 считается, что нечеткая точка лежит на кривой, а если $M_i\,(a)$ близка к нулю – «точка Q_i кривой не принадлежит».

Для *f-регрессии* исследовано два способа формального определения нечетких точек:

1. С использованием *T*-нормы «произведение». Предположим, что разбросы всех входных и выходной переменной независимы. Тогда функцию принадлежности нечеткой точки $Q_i = (X_i,Y_i)$ можно записать в виде

$$\mu_{\boldsymbol{Q}_i}\left(\boldsymbol{x}, y\right) = \mu_{X_{1i}}(x_1)\cdot\ldots\cdot\mu_{X_{ni}}(x_n)\cdot\mu_{Y_i}(y)\,. \qquad (4)$$

Для линейной модели метод неопределенных множителей Лагранжа позволяет, благодаря выбранному виду (2), вывести выражение для *меры соответствия* $M_i(a)$ между i-й нечеткой точкой и произвольной параметрической прямой $y = f(x,a)$ в явном виде как функцию параметров модели a:

$$M_i(\mathbf{a}) = L\left(\frac{\left|d_i\,(\mathbf{a})\right|}{\left[s_i\,(\mathbf{a})\right]^{\frac{m-1}{m}}}\right), \qquad (5)$$

где

$$d_i(\boldsymbol{a}) = a_0 + \sum_{j=1}^{n} a_j\cdot\bar{x}_{ji} - \bar{y}_i\,, \qquad (6)$$

$$s_i(\boldsymbol{a}) = \alpha_i^{\frac{m}{m-1}} + \sum_{j=1}^{n}\left|a_j\cdot\beta_{ji}\right|^{\frac{m}{m-1}}, \qquad (7)$$

причем точка (x_i^*, y_i^*) на прямой $y = f(x,a)$, в которой достигается данная *мера соответствия* между i-й точкой и линейной моделью с параметрами a, определяется следующим выражением:

$$x_{ji}^*(\boldsymbol{a}) = \bar{x}_{ji} - \operatorname{sgn}(a_j)\cdot\left|a_j\right|^{\frac{1}{m-1}}\cdot\beta_{ji}^{\frac{m}{m-1}}\cdot\frac{d_i\,(\boldsymbol{a})}{s_i\,(\boldsymbol{a})}\,,$$

$$y_i^* = f(\boldsymbol{x}_i^*, \boldsymbol{a})\,. \qquad (8)$$

2. Эллиптические нечеткие точки. Данный способ представления исходных данных используется в случае наличия корреляции (взаимной зависимости) неопределенностей различных переменных. Поскольку для инженерной задачи, описанной в настоящей статье, такая ситуация нетипична, мы отсылаем читателя к статье [7] за дальнейшей информацией об эллиптических точках. Пример решения задачи методом *f-регрессии* с использованием данного вида нечетких точек рассматривается в [8].

Мера соответствия между ансамблем данных и параметрической кривой

Описанная выше нечеткая *мера соответствия* для единичной точки может быть распространена на весь ансамбль из N точек для получения меры соответствия параметрической модели ансамблю. В работе [5] впервые предложено семейство агрегирующих операторов, имеющих смысл **нечетких мер истинности** (fuzzy plausibility measure) (Закрепившийся в англоязычной традиции термин “plausibility” дословно переводится, как «правдоподобие, вероятность». По причине закрепления этих терминов в русскоязычной традиции за понятиями теории вероятности, здесь “plausibility” используется в значении «истинность»). Для метода *f-регрессии* оптимальным вектором параметров модели является такой вектор ***a****, для которого достигается максимум нечеткой меры истинности.

Концепцию нечеткой истинности проще всего проиллюстрировать следующим примером. Предположим, что у нас имеется нечеткая точка, заданная функцией принадлежности $\mu_{\boldsymbol{Q}_i}\left(\boldsymbol{x}, y\right)$ (4). Тогда для заданной модели $y = f(x,a)$ и произвольного вектора параметров ***a***, **мерой максимальной истинности** утверждения «нечеткая точка Q_i принадлежит кривой f с параметрами a» будет *мера соответствия* $M_i(\boldsymbol{a})$ (5).

Таким образом, описанные в работах [5], [9], [8] агрегирующие операторы имеют смысл следующих мер истинности $MP(\boldsymbol{a})$ [10]:

- **Среднее геометрическое** дает оценку истинности утверждения «модель с параметрами a проходит через все точки ансамбля»:

$$MG(\boldsymbol{a}) = \prod_i M_i^{w_i}(\boldsymbol{a}), \qquad (9)$$

где w_i – веса экспериментальных точек.

Геометрическое среднее аналогично дизъюнкции индивидуальных мер соответствия нечетких точек модели, и в развернутом виде может быть представлено как «модель с параметрами $\boldsymbol{a}$ проходит через точку 1 **И** через точку 2 **И** … **И** через точку N». В этом смысле критерий (9) близок по смыслу к критерию наименьших квадратов; фактически, как показано в [10], при определенном сочетании параметров они математически эквивалентны, то есть в некотором смысле метод наименьших квадратов является частным случаем метода *f-регрессии* с мерой среднее геометрическое.

• **Среднее арифметическое** дает оценку истинности утверждения «модель с параметрами a проходит через **большинство** точек ансамбля»:

$$MA(\boldsymbol{a}) = \sum_i w_i M_i(\boldsymbol{a}). \qquad (10)$$

Практический эффект применения данной меры истинности заключается в **автоматическом отсеивании выпадающих наблюдений**, поскольку по определению выбросы – это точки, отличающиеся от большинства.

• **Среднее квадратическое** дает оценку истинности утверждения «модель с параметрами $\boldsymbol{a}$ проходит через **группу** точек ансамбля»:

$$MQ(\boldsymbol{a}) = \sqrt{\sum_i w_i M_i^2(\boldsymbol{a})}. \qquad (11)$$

Данная мера истинности усиливает эффект (11) за счет того, что она отдает предпочтение «элитным» группам точек.

Для метода *f-регрессии* выбранная нечеткая мера истинности $MP(\boldsymbol{a})$ определяет смысл «наилучшего» вектора параметров модели $\boldsymbol{a}_{MP}$ следующим образом.

Определение 5. *Максимально истинные оценки* параметров – это такой вектор параметров $\boldsymbol{a}_{MP}$, при котором для заданного ансамбля экспериментальных точек достигается максимум выбранной нечеткой меры истинности

$$\boldsymbol{a}_{MP} = \arg\max_{a \in \mathbf{R}^p} MP(\boldsymbol{a}). \qquad (12)$$

Поскольку параметры нечетких чисел, например, параметр формы m, влияют на получаемые оценки, в нотации принято указывать их значения там, где это необходимо, в следующем виде: $\boldsymbol{a}_{MQ}^{m=4}$.

В идеальном случае, когда все экспериментальные точки лежат на выбранной модели, все три меры дают одинаковую оценку параметров $\boldsymbol{a}_{MG} = \boldsymbol{a}_{MA} = \boldsymbol{a}_{MQ}$. На практике, однако, в случае неоднородностей экспериментальных данных и особенно присутствия «выпадающих» наблюдений, их максимально истинные оценки будут различаться. В общем случае, максимизация разных MP дает различные уникальные оценки параметров модели, являющиеся «максимально истинными оценками» в их специальном смысле.

Нелинейный метод f-регрессии

В общем случае невозможно получить аналитические выражения для меры соответствия (5) между нечеткой точкой и произвольной нелинейной кривой. Однако, в случае, если модель может быть приведена к линейному по параметрам виду, линейный регрессионный анализ может быть проведен с преобразованными нечеткими числами, представляющими исходные данные. Преобразования L-чисел, основанные на разложении преобразующих функций в ряд Тейлора первого порядка, подробно описаны в [8].

2 Кластерная *f*-регрессия

Впервые подход к решению задачи классификации при помощи *f*-регрессии был предложен в работе [8]. В настоящей работе дается обобщение подхода к регрессионному анализу экспериментальных данных, представляющих несколько классов, каждый из которых описывается собственной линейной моделью с параметрами $\boldsymbol{a}^{(1)}$, $\boldsymbol{a}^{(2)}$ и т.д. Мера истинности нечеткого утверждения «i-я точка принадлежит Классу 1 **ИЛИ** Классу 2 **ИЛИ** …», или, для краткости, «i-я точка принадлежит **любому** классу», обозначаемая $M_i^C(\ldots)$, является расширением обычной меры соответствия (5) и задается следующим выражением:

$$M_i^C\left(a^{(1)},\ldots,a^{(k)}\right) = M_i\left(a^{(1)}\right) \tilde{+} \ldots \tilde{+} M_i\left(a^{(k)}\right), \qquad (13)$$

где k – общее количество классов, выбираемое экспериментатором или путем автоматической процедуры (в настоящей работе на рассматривается);

$\tilde{+}$ – S-норма (T-конорма) «**алгебраическая сумма**».

Любой из описанных выше агрегирующих критериев истинности (9)-(11) может использоваться в сочетании с кластерной мерой соответствия $M_i^C(\ldots)$.

В **разделе 3** рассматривается пример использования кластерного регрессионного анализа и приводятся практические соображение относительно использования данного подхода.

3 Анализ данных испытания скважины

В качестве примера рассматриваются данные восстановления давления на скважине с эффектом послепритока и скин-фактором, вскрывающей пласт с двойной моделью пористости (Пример 2 в работе Бурде и др. [10]). Данные производной забойного давления обрабатываются методом кластерной *f*-регрессии, описанным в **разделе 2**. В силу ограниченности места содержательная интерпретация полученных данных и сопоставление с результатами оценки параметров, полученными Бурде и др., не приводится, поэтому пример преследуют в основном демонстрационную цель.

Применительно к задаче анализа кривой восстановления давления можно констатировать, что интерпретация КВД экспертом вручную основана на следующих особенностях человеческого интеллекта:

1. Способность визуально выделять линейные участки в экспериментальных данных, при этом отфильтровывая отдельные сильно выбивающиеся из общей закономерности точки («выбросы» или «выпадающие наблюдения»).

2. Слияние скоплений точек. Невозможно и нет необходимости определять, сколько конкретно точек находится в скоплении – 20, 50 или 100, поэтому они воспринимаются и обрабатываются как единая совокупность.

3. Отнесение воспринимаемого образа кривой к одному из известных типичных КВД для разных моделей скважины, пласта и границ. В этом процессе важную роль играет теоретическая подготовка и практический опыт.

Перечисленные способности человеческого интеллекта находят следующее отражение в подходе кластерного регрессионного анализа:

1. Данные аппроксимируются линейными моделями. В качестве критерия максимальной истинности выбрана мера «среднее квадратическое», имманентным свойством которого является способность идентифицировать модель, описывающую группу нечетких точек (следовательно, отфильтровывать «выпадающие» наблюдения).

2. Значимость каждой экспериментальной точки для визуального и, следовательно, регрессионного анализа тем меньше, чем больше «кучность» точек в ее окрестности. Формализация воспринимаемой «плотности» δ_i экспериментальных данных в окрестности i-й точки в логарифмических координатах, определяющей нормализованный вес точки w_i, производится следующим образом:

$$\delta_i = \sum_{j=1}^{N} L\left(\frac{1}{\omega} \cdot \left|\log_{10}\left(\frac{t_D}{C_D}\right)_i - \log_{10}\left(\frac{t_D}{C_D}\right)_j\right|\right),$$

$$w_i = \frac{\delta_i^{-1}}{\sum_{j=1}^{N} \delta_j^{-1}}. \qquad (16)$$

Здесь настраиваемый параметр ω определяет ширину «окна», внутри которого точки i и j считаются находящимися «рядом» друг с другом на логарифмической оси безразмерного времени. Поскольку понятие «рядом» также является размытым (нечетким), в его определении используется уже знакомая L-функция (2).

3. Данная способность не является свойством метода как такового, поскольку метод *f-регрессии* не привязан к конкретной предметной области. Тем не менее, модели типичных КВД и эмпирические знания эксперта могут быть заложены в предлагаемый метод в виде ограничений на значения параметров кусочно-линейной модели – например, начальный период с $a \approx (0; 1)$ или бесконечно-радиальный период с $a \approx (0{,}5; 0)$. Здесь же мы ограничимся тем, что подчеркнем: нечеткий критерий максимальной истинности **одновременно** является индикатором **адекватности** выбранной модели экспериментальным данным и легко интерпретируемым показателем **качества** модели. Таким образом, результатом кусочно-линейной аппроксимации является оценка адекватности выдвинутых предположений о характере модели с учетом имеющихся априорных сведений о неопределенности исходных данных.

При анализе были сделаны следующие предположения относительно неопределенности входной переменной $\log_{10}\left(\frac{t_D}{C_D}\right)$ (логарифм безразмерного времени с момента остановки скважины):

• Шаг дискретизации по времени составляет ~1,85 секунды. Поскольку рассматривается разность между текущим временем и неизвестным моментом остановки скважины, базовый разброс времени составляет ~0,92 секунды.

• Поскольку данные в таблице представлены числами, округленными до 5 значащих цифр, имеет место ошибка округления, зависящая от абсолютного значения времени.

• Пересчет времени в безразмерное время осуществляется по формулам, в которых присутствует значительно количество параметров модели, содержащих неопределенности, в том числе в силу непостоянства этих параметров во времени: проницаемость, пористость, мощность пласта, динамическая вязкость и т.д. [1]. Однако, в силу принятого в [10] методического подхода и имеющего место факта предварительной обработки данных, данный компонент было решено не учитывать.

• Логарифмирование Δt приводит к преобразованию функции принадлежности исходного нечеткого числа. В общем случае результирующая функция принадлежности должна определяться применением принципа обобщения Заде, однако, в методе *f-регрессии* принят упрощенный подход (см. раздел «1.3 Нелинейный метод *f-регрессии*»).

При анализе были сделаны следующие предположения относительно неопределенности выходной переменной $\log_{10} p'_D$ (логарифм производной безразмерной разности давлений по логарифмическому безразмерному эффективному времени):

• Погрешность измерения забойного давления равна 0,01 бар.

• Как и в случае измерений времени, имеет место ошибка округления.

• Случайные флуктуации давления, происходящие в силу множества природных факторов, очевидным образом сведены к минимуму за счет предварительной обработки данных. Косвенным подтверждением этому служит то, что представленные в [10] данные изменения разности давления носят строго монотонно возрастающий характер.

ISSN 1993—9868. Нафтогазова енергетика. 2013. № 2(20)

• Пересчет разности давления в безразмерные величины, по сравнению с безразмерным временем, осуществляется с использованием двух дополнительных параметров: приведенного дебита в момент остановки скважины и коэффициента сжимаемости. Неопределенность указанных компонентов не учитывается.

• Формула для пересчета времени Δt в эффективное время $t_e = \dfrac{t_p \cdot \Delta t}{t_p + \Delta t}$ является источником неопределенности в первую очередь в силу сделанного при расчете приведенного времени t_p предположения о характере фильтрационного течения на предшествующем остановке скважины периоде. Поэтому в целях анализа было принято допущение о существовании неопределенности методического происхождения в значении эффективного времени, не зависящей прочих параметров и составляющей 1% от его абсолютного значения. Данное допущение является наименее обоснованным и служит предметом для открытой дискуссии. Отметим, что приведенное время оказывает наиболее существенное влияние именно в позднее время ($t_D / C_D > 5000$).

• Методика расчета производной давления по логарифмическому эффективному времени, предложенная в [12], определяет производную как средневзвешенное значения производных на соседних участках и содержит большое количество арифметических операций. Применение последовательности этих операций к нечетким числам приводит к лавинообразному разрастанию их параметров разброса, составляющему в среднем для ансамбля 163% от среднего получаемых значений. Поэтому в целях определения разброса производной давления была применена простая, невзвешенная формула, дающая в среднем для ансамбля 15% от величины среднего. Мы исходим из предположения, что данная замена не искажает характер неопределенности производной, поскольку применение средневзвешенной формулы преследует цель снижения неопределенности значения производной.

При выборе параметра формы нечетких чисел мы руководствовались следующими соображениями:

• Измеренные значения времени и давления содержат неопределенности интервального характера, такие как шаг дискретизации и ошибка округления, которые моделируются с применением параметра формы $m > 10$.

• Измерения также содержат неопределенности, связанные с перемножением нечетко определенных параметров, возникающие в момент перехода к безразмерным величинам. Проводя вполне уместные в данном случае аналогии с теоремами теории вероятностей, можно предположить логнормальную форму функции принадлежности безразмерных давления и времени. Далее, поскольку в целях анализа кривых происходит логарифмирование этих переменных, результирующая форма неопределенности будет иметь вид, математически эквивалентный нормальной плотности вероятности, который моделируется при помощи параметра формы $m = 2$.

• Результирующий параметр формы нечетких данных следует выбирать в диапазоне $2 < m < 10$. Эмпирическим путем установлено, что наиболее адекватные результаты получаются при использовании параметра формы $m = 4$.

Результаты аппроксимации 151 экспериментальной точки 4 линейными фрагментами показаны на рис. 1. Данная аппроксимация подразумевает решение задачи без ограничений в восьмимерном пространстве параметров. Решение этой задачи дает оценку адекватности (истинности) полученной совокупности линейных моделей и максимально истинные оценки (МИО) параметров кластерной *f*-регрессии, представленные в Таблице 1. Ниже приводится комментарий относительно полученных результатов.

Раннее время ($t_D / C_D < 12$) представлено *(1)* эффектом **первичного накопления** нефти в стволе скважины в результате послепритока, описываемого *Классом 1* (значение $\mathbf{a}_1^{(1)} \pm \mathbf{1}$ дает повод предположить, что момент остановки скважины определен неточно), *(2)* дугой максимального значения производной, *(3)* периодом **снижения производной**, описываемым *Классом 2* (см. Примечание в конце раздела) и *(4)* псевдорадиальным переходным процессом (в данном примере отсутствует).

Среднее время ($t_D / C_D < 5000$) демонстрирует характерный для резервуаров **двойной пористости** прогиб производной со значениями $p'_D < 0{,}5$. Параметры *Класса 3* позволяют более точно оценить параметры двойной пористости.

Позднее время ($t_D / C_D > 5000$) показывает бесконечно-радиальное течение гомогенизированной пластовой системы с теоретическим $p'_D = 0{,}5$ (постоянное значение). Отрицательный угол наклона *Класса 4* $\mathbf{a}_1^{(4)} = \mathbf{-0{,}04}$ дает повод предположить, что давление в системе постепенно приближается к пластовому.

Примечание: При рассмотрении результатов анализа можно сделать два наблюдения. Во-первых, тенденция к обобщению зависимостей на весь ансамбль экспериментальных данных может привести к искажению оценок (см. пунктирное продолжение модели *Класса 4* вызывает смещение модели *Класса 2*). Во-вторых, бросается в глаза отсутствие переходов между фрагментами, на самом деле аппроксимирующими непрерывный по природе процесс. Возможно, расширения для кусочно-линейной регрессии или модели Сугено позволят добиться лучшей интерпретируемости результатов.

Таблица 1 – Результаты кластерного *f*-регрессионного анализа данных восстановления давления

Классы	МИО $a_{MQ}^{(i)}$ $p'_D = 10^{a_0}\left[\dfrac{t_D}{C_D}\right]^{a_1}$		Количество точек, принадлежащих модели $(M_i(a) > 0.9)$		Количество "выпадающих" наблюдений $(M_i(a) < 0.1)$		$MQ(a)$
	a_0	a_1	Кол-во	Взвеш.	Кол-во	Взвеш.	
Класс1	–0.1983	0.6797	6	17.5	144	130.1	0.3413
Класс2	0.2010	–0.2617	21	23.5	113	109.6	0.4440
Класс3	0.3192	–0.5839	22	47.5	126	98.2	0.5639
Класс4	–0.1348	–0.0403	48	18.3	97	128.9	0.3539
Все классы			94	94.1	37	38.3	0.8150

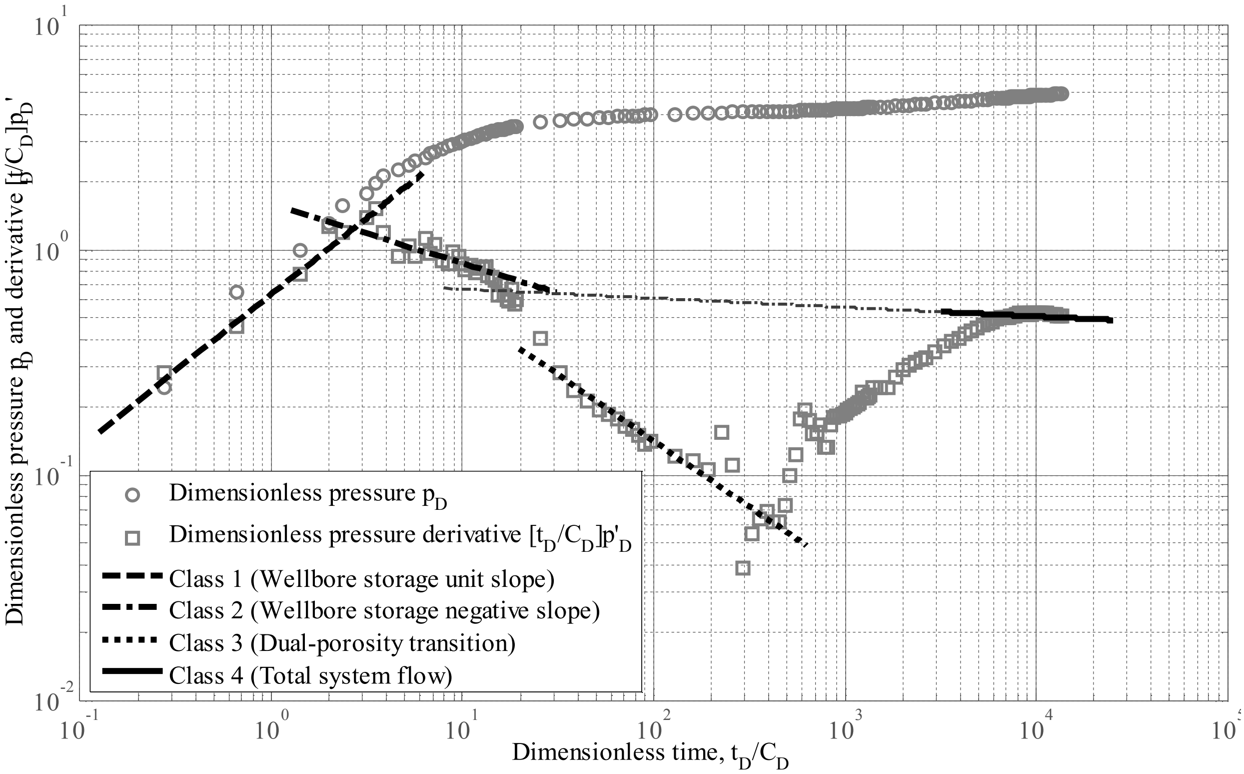

Рисунок 1 – Аппроксимация данных КВД в резервуаре двойной пористости в билогарифмических координатах

Вывод

В работе продемонстрирована эффективность подхода к кластерному регрессионному анализу данных испытаний скважин методом *f-регрессии,* основанным на мере соответствия нечеткой точки и параметрической модели. Показано, каким образом подход эксперта формализуется при помощи аппарата теории нечетких множеств в виде нечеткого критерия максимальной истинности на основании неопределенностей, заданных в явном виде. Формулировка критерия подразумевает один из следующих смыслов оптимальности получаемых оценок параметров: прохождение модели через все точки одновременно, через большинство точек, или через группу точек. Показано, что нечеткая меры истинности «среднее геометрическое» отфильтровывает точки, «выпадающие» из ансамбля.

Подход позволяет построить автоматическую процедуру кластеризации данных и идентификации параметров, которая может быть адаптирована для решения более специфических задач анализа данных испытаний скважин за счет добавления экспертных знаний о модели скважины, пласта или границ, а также для решения широкого спектра других задач анализа.

Литература

1 Bourdet D.P. "Well Test Analysis: the Use of Advanced Interpretation Models". 1st ed. The Netherlands: Elsevier, 2002.

2 Zadeh L.A. "Generalized theory of uncertainty (GTU) – principal concepts and ideas" // Computational Statistics & Data Analysis. 2006. No. 51. pp. 15-46.

ISSN 1993—9868. Нафтогазова енергетика. 2013. № 2(20)

3 Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Поспелова Д.А.. – М.: Наука, 1986.

4 Tanaka H., Uejima S., and Asai K. "Linear regression analysis with fuzzy model" // IEEE Transaction on System, Man and Cybernetics. 1982. No. 12. pp. 903-907.

5 Kalinina E., Wagenknecht M. "Fuzzy regression analysis and application to a crisp model" // Proceedings of the 8th Zittau Fuzzy Colloquium. Sep 2000. pp. 9-18.

6 Bandemer H., Gottwald S. "Einfuehrung in Fuzzy Methoden" (на немецком). Berlin: Akademie-Verlag, 1989.

7 Celminš A., "Least squares model fitting to fuzzy data". Fuzzy Sets and Systems, No. 22, 1987. pp. 245-269.

8 Izyumov B. "Application of f-regression method to fuzzy classification problem" // Proceedings of the EUSFLAT Conference, Zittau. 2003. pp. 761-766.

9 Izyumov B., Kalinina E., and Wagenknecht M. "Software tools for regression analysis of fuzzy data" // Proceedings of 9th Zittau Fuzzy Colloquium. 2001. pp. 221-229.

10 Bourdet D.P., Ayoub J.A., and Pirard Y.M., "Use of Pressure Derivative in Well Test Interpretation (SPE 12777)", SPEFE, 1989. pp. 293-302.

Стаття надійшла до редакційної колегії
27.10.13
Рекомендована до друку
професором ***Семенцовим Г.Н.***
(ІФНТУНГ, м. Івано-Франківськ)